



## CERTIFICATE OF EXPRESS MAILING

Express Mail Mailing Label Number ED 144345983 US

Date of Deposit 3 December 2004

Pursuant to 37 C.F.R. § 1.10, I certify that I am personally depositing the following paper(s) or fee(s) with the "Express Mail Post Office to Addressee" service of the United States Postal Service on the above date in a sealed envelope (a) having the above-numbered Express Mail label and sufficient postage affixed, and (b) addressed to: Mail Stop Appeal Brief-Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA, 22313-1450

Stamped, self-addressed postcard for USPTO receipt stamp (1 card)

Fee Transmittal Form (Form PTO/SB/17) (1 sheet)

Credit Card Payment Form (Form PTO-2038) (1 sheet)

Appeal Brief (14 sheets)

Appendix A (4 sheets)

Appendix B (6 sheets)

Application No.: 09/867,803

Art Unit: 2177

Confirmation No.: 6606

Examiner: Khanh B. Pham

Application Filing Date: 31 May 2001

Inventor: CHOI et al.

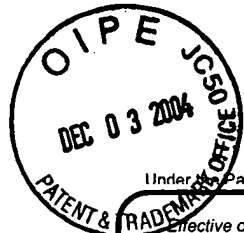
Document Submission Date: 3 December 2004

Docket: 1005-006

3 December 2004  
Date

Eden Brown  
Name of Person Making Deposit

Eden Brown  
Signature of Person Making Deposit



PTO/SB/17 (11-04)

Approved for use through 07/31/2006. OMB 0651-0032

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Effective on 10/01/2004. Patent fees are subject to annual revision.

# FEE TRANSMITTAL

## For FY 2005

☒ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$ ) 225.00

**Complete if Known**

Application Number	09/867,803
Filing Date	31 May 2001
First Named Inventor	Lawrence J. Choi
Examiner Name	Khanh B. Pham
Art Unit	2177
Attorney Docket No.	1005-006

**METHOD OF PAYMENT** (check all that apply)☐ Check ☒ Credit Card ☐ Money Order☐ Deposit Account ☐ NoneDeposit Account Number  
50-2504Deposit Account Name  
Michael N. Haynes

The Director is hereby authorized to: (check all that apply)

- ☐ Charge fee(s) indicated below
- ☐ Charge fee(s) indicated below, **except for the filing fee**
- ☒ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17
- ☒ Credit any overpayments

to the above-identified deposit account.

☐ Other (please identify):**WARNING:** Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**FEE CALCULATION****1. BASIC FILING FEE**

Fee Description	Fee (\$)	Small Entity Fee (\$)	Fee Paid(\$)
Utility Filing Fee	790	395	
Design Filing Fee	350	175	
Plant Filing Fee	550	275	
Reissue Filing Fee	790	395	
Provisional Filing Fee	160	80	

Subtotal (1) \$ 0.00

**FEE CALCULATION** (continued)**2. EXTRA CLAIM FEES**

Fee Description	Fee (\$)	Small Entity Fee (\$)
Each claim over 20	18	9
Each independent claim over 3	88	44
Multiple dependent claims	300	150
For Reissues, each claim over 20 and more than in the original patent	18	9
For Reissues, each independent claim more than in the original patent	88	44

**Total Claims**      **Extra Claims**      **Fee (\$)**      **Fee Paid (\$)**- 20 or HP = \_\_\_\_\_ x \_\_\_\_\_ = \_\_\_\_\_  
HP = highest number of total claims paid for, if greater than 20**Indep. Claims**      **Extra Claims**      **Fee (\$)**      **Fee Paid (\$)**- 3 or HP = \_\_\_\_\_ x \_\_\_\_\_ = \_\_\_\_\_  
HP = highest number of independent claims paid for, if greater than 3**Multiple Dependent Claims**      **Fee (\$)**      **Fee Paid (\$)**

Subtotal (2) \$ 0.00

**3. OTHER FEES**

Fee Description	Fee (\$)	Small Entity Fee (\$)	Fee Paid(\$)
1-month extension of time	110	55	55.00
2-month extension of time	430	215	
3-month extension of time	980	490	
4-month extension of time	1,530	765	
5-month extension of time	2,080	1,040	
Information disclosure stmt. fee	180	180	
37 CFR 1.17(q) processing fee	50	50	
Non-English specification	130	130	
Notice of Appeal	340	170	
Filing a brief in support of appeal	340	170	170.00
Request for oral hearing	300	150	
Other:			

Subtotal (3) \$ 225.00

**SUBMITTED BY**

Signature		Registration No. (Attorney/Agent) 40,014	Telephone 434-972-9988
Name (Print/Type)	Michael N. Haynes		Date 3 December 2004

This collection of information is required by 37 CFR 1.136. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 30 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.



12.06.04

AE/21778  
PATENT  
EPW

Serial No. 09/867,803

Attorney Docket No. 1005-006

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) : Lawrence J. Choi et al.  
Serial No. : 09/867,803  
Filed : 31 May 2001  
For : METHOD AND SYSTEM FOR CLUSTERING  
OPTIMIZATION AND APPLICATIONS  
Art Unit : 2177  
Examiner : Khanh B. Pham

**Mail Stop Appeal Brief-Patents**

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

12/08/2004 HALI11 00000105 09867803

01 FC:2402  
02 FC:2251

170.00 OP  
55.00 OP

**APPEAL BRIEF**

Sir:

The Applicants respectfully submit this appeal brief in response to the Office Action of 3 May 2004 finally rejecting each of the pending claims 1, 2, and 5-8. This Appeal Brief is in furtherance of the Notice of Appeal filed 3 September 2004. In addition to paying the requisite fee for an Extension of Time under 37 C.F.R. §1.136(a), Applicant petitions to extend the due date for this Appeal Brief to 3 December 2004.

**I. REAL PARTY IN INTEREST**

The real party in interest is Rosetta Marketing Strategies Group, a corporation having a place of business at 103 Carnegie Center, Suite 202, Princeton, New Jersey 08540. *See* Assignment (recorded 31 May 2001 at reel 011881, frame 0782).

**II. RELATED APPEALS AND INTERFERENCES**

Also under appeal is the final rejection of claims of United States Patent Application 09/867,582.

**III. STATUS OF CLAIMS**

Claims 1-8 are pending in this application. Claims 1, 2, and 5-8 have been finally rejected. Claims 3 and 4 have been finally objected to. Claims 1-8 are the subject of this appeal.

**IV. STATUS OF AMENDMENTS**

A Reply to the final rejection was filed on 2 July 2004, that Reply containing amendments to claim 3. That amendment to claim 3 was entered in an Advisory Action dated 30 August 2004.

**V. SUMMARY OF CLAIMED SUBJECT MATTER**

**Independent Claim 1**

Claim 1 defines a computer-assisted method (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5) for evaluating a cluster assignment for an observation (see at least FIG. 4, including elements 4010 and 4090; page 50, lines 24-26). The method comprises the activities of:

for each of a plurality of observations (see at least FIG. 4, element 4010; page 50, lines 24-26), obtaining a data set containing no more than one proxy value for each of a plurality of variables (see at least FIG. 4, element 4010; page 50, lines 24-26), each variable having a plurality of possible values (see at least FIG. 4, element 4010; page 50, lines 24-26), the data set also containing a cluster assignment for the observation (see at least FIG. 4, including element 4010; page 50, lines 24-26), the cluster assignment identifying one cluster from a plurality of clusters (see at least FIG. 4, elements 4010 and 4020; page 50, lines 24 through page 51, line 5);

for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables (see at least FIG. 4, elements 4080; page 52, lines 8-10); and

outputting the percent for each observation (see at least FIG. 4, elements 4085; page 52, lines 10-11).

## **Independent Claim 2**

Claim 2 defines a computer-assisted method (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5) for evaluating a cluster assignment for an observation (see at least FIG. 4, including elements 4010 and 4090; page 50, lines 24-26), comprising the activities of:

for each of a plurality of observations (see at least FIG. 4, element 4010; page 50, lines 24-26), obtaining a data set containing no more than one proxy value for each of a plurality of variables (see at least FIG. 4, element 4010; page 50, lines 24-26), each variable having a plurality of possible values (see at least FIG. 4, element 4010; page 50, lines 24-26), the data set also containing a cluster assignment for the observation (see at least FIG. 4, element 4010; page 50, lines 24-26);

for each observation from the plurality of observations (see at least FIG. 4, element 4020; page 50, lines 24 through page 51, line 5), estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters (see at least FIG. 4, element 4020; page 50, lines 24 through page 51, line 5); and  
outputting each purposeful probability (see at least FIG. 4, element 4025; page 51, lines 13-14).

### **Dependent Claim 3**

In addition to the subject matter of claim 1, claim 3 defines:

for each observation from the plurality of observations in each cluster from the plurality of clusters (see at least FIG. 4, element 4030; page 51, lines 14-18), calculating a serendipity probability for each possible value (see at least FIG. 4, element 4030; page 51, lines 14-18), the serendipity probability is a measure of a probability that an observation in a particular cluster will be randomly associated with any one of the plurality of possible values for a particular variable (see at least page 26, line 21 through page 27, line 10);

for each observation from the plurality of observations, calculating a ratio of the purposeful probability to the serendipity probability (see at least FIG. 4, element 4035; page 51, lines 19-20);

for each observation from the plurality of observations, calculating a logarithm of the ratio to obtain composition analysis score (see at least FIG. 4, element 4040; page 51, lines 20-22); and

outputting the composition analysis scores for each observation in each cluster (see at least FIG. 4, element 4045; page 51, lines 22-23).

**Dependent Claim 4**

In addition to the subject matter of claim 1, claim 4 defines:

for each observation from the plurality of observations, assuming that before the observation can be made, the observation has an equal probability of being in any identified cluster from the plurality of clusters (see at least FIG. 4, element 4050; page 51, lines 24-25);

for each observation from the plurality of observations, assuming that the purposeful probabilities are true (see at least FIG. 4, element 4055; page 52, lines 1-4);

for each observation from the plurality of observations, using Bayes' Theorem to calculate a Bayes probability that a particular observation can be in each cluster conditional upon the observation's proxy value to each variable (see at least FIG. 4, element 4060; page 52, lines 4-6);

outputting the Bayes probability that each observation can be in each cluster (see at least FIG. 4, elements 4065; page 52, lines 6-7).

**Independent Claim 5**

Claim 5 defines a computer-readable medium containing instructions for activities (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5). Those activities comprise:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters (see at least FIG. 4, elements 4010 and 4020; page 50, lines 24 through page 51, line 5);

for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables (see at least FIG. 4, elements 4080;

page 52, lines 8-10); and

outputting the percent for each observation (see at least FIG. 4, elements 4085; page 52, lines 10-11).

### **Independent Claim 6**

Claim 6 defines an apparatus (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5) for evaluating a cluster assignment for an observation (see at least FIG. 4, including elements 4010 and 4090; page 50, lines 24-26), comprising:

for each of a plurality of observations, means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and lines 10-12) obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters (see at least FIG. 4, elements 4010 and 4020; page 50, lines 24 through page 51, line 5);

for each observation from the plurality of observations, means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and lines 10-12) calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables (see at least FIG. 4, elements 4080; page 52, lines 8-10); and

means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and line 26 through page 61, line 2) outputting the percent for each observation (see at least FIG. 4, elements 4085; page 52, lines 10-11).



**Independent Claim 7**

Claim 7 defines a computer-readable medium containing instructions for activities (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5). These activities comprise:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation (see at least FIG. 4, elements 4010 and 4020; page 50, lines 24 through page 51, line 5);

for each observation from the plurality of observations, estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters (see at least FIG. 4, element 4020; page 50, lines 24 through page 51, line 5); and

outputting each purposeful probability (see at least FIG. 4, element 4025; page 51, lines 13-14).

**Independent Claim 8**

Claim 8 defines an apparatus (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5) for evaluating a cluster assignment for an observation (see at least FIG. 4, including elements 4010 and 4090; page 50, lines 24-26), comprising:

for each of a plurality of observations, means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and lines 10-12) obtaining a data set containing no more than one proxy value for each of a plurality of variables, each

variable having a plurality of possible values, the data set also containing a cluster assignment for the observation (see at least FIG. 4, elements 4010 and 4020; page 50, lines 24 through page 51, line 5);

for each observation from the plurality of observations, means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and lines 10-12) estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters (see at least FIG. 4, element 4020; page 50, lines 24 through page 51, line 5); and

means for (see at least FIG. 4; page 50, line 21 through page 52, line 13; FIG. 9, elements 9100-9500; page 59, lines 21-27, and page 60, line 1 through page 61, line 5, and more particularly page 60, line 1-3 and line 26 through page 61, line 2) outputting each purposeful probability (see at least FIG. 4, element 4025; page 51, lines 13-14).

## **VI. GROUNDS OF REJECTION**

Claim 3 was objected to because of an informality.

Claims 3 and 4 were objected to as being dependent upon a rejected base claim, but were indicated as allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Claims 1-2 and 5-8 were rejected as anticipated under 35 U.S.C. §102(e). In support of these rejections, Fayyad (U.S. Patent No. 6,633,882) was cited. These rejections are respectfully traversed.

## VII. ARGUMENT

### I. Generally

Fayyad fails to establish a *prima facie* case of anticipation. See MPEP 2131. To anticipate expressly, the “invention must have been known to the art in the detail of the claim; that is, all of the elements and limitations of the claim must be shown in a single prior art reference, arranged as in the claim”. *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383, 58 USPQ2d 1286, 1291 (Fed. Cir. 2001). The single reference must describe the claimed subject matter “with sufficient clarity and detail to establish that the subject matter existed in the prior art and that its existence was recognized by persons of ordinary skill in the field of the invention”. *Crown Operations Int’l, LTD v. Solutia Inc.*, 289 F.3d 1367, 1375, 62 USPQ2d 1917, 1921 (Fed. Cir. 2002). Moreover, the prior art reference must be sufficient to enable one with ordinary skill in the art to practice the claimed invention. *In re Borst*, 345 F.2d 851, 855, 145 USPQ 554, 557 (C.C.P.A. 1965), *cert. denied*, 382 U.S. 973 (1966); *Amgen, Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1354 (Fed. Cir. Jan. 6, 2003) (“A claimed invention cannot be anticipated by a prior art reference if the allegedly anticipatory disclosures cited as prior art are not enabled”).

Fayyad allegedly recites an “[a]pparatus and method for use in querying a database containing data records. The database is characterized by a compression scheme to provide data clustering information. In accordance with a exemplary embodiment of the invention a functional representation of data clustering is a **Gaussian** and the queries are performing by integrating the Gaussian corresponding to each of the data clusters over the ranges to determine the sum or the count of data records from the database that fall within the selected ranges. The process chooses a value for the cluster number K. The cluster model is next broken up into areas (tiles) based on user defined parameters. Data from the database is then classified based on the tiling information. A sorted version of the classified data, ordered by cluster number and then by the tile number

within the cluster is generated. This data is then evaluated to test the sufficiency of the model created during the clustering.” See Abstract.

Also, Fayyad allegedly recites that an “exemplary embodiment of the invention starts with one cluster having an arbitrarily assigned **mean** or centroid for each dimension of data within the database. This one cluster is selected without accessing the data in the database. Using this starting point data is retrieved or read from the database and a cluster model is built from the data using the single starting cluster.” See Fayyad’s Summary of the Invention, col. 3, lines 50-54.

Fayyad allegedly recites that “[i]n accordance with the exemplary embodiment of the present invention, the number of clusters K is determined by an iterative process that is summarized in FIG. 7.... Once this initialization step 100 is performed, the scalable clustering procedure that uses the E-M, **K-Means** or other suitable clustering process is performed until a stopping point 140 is reached.” See col. 8, lines 1-9.

Further, Fayyad allegedly recites that “[w]hen used with the scalable clustering process of FIG. 3, the initialization process of FIG. 7 selects a starting cluster number, K and determines a starting point for the **means** or centroids of the K clusters”. Col. 8, lines 10-13.

Fayyad allegedly recites that the “FIG. 7 process operates by identifying areas (partitions) of multidimensional space that have a higher or lower **density** of points than is predicted by a current cluster model. The current model is then further refined by growing new clusters in the areas of low or high density so that the new model better fits the data. This improves the density estimation of the clustering and hence improves the accuracy of queries answered using the clustering model.” Col. 10, lines 16-24.

Finally, Fayyad allegedly recites that a “number of accuracy parameters are used to control the clustering initialization process of FIG. 7.” See col. 9, lines 64-65. “An accuracy parameter (TileAccuracy)” “is the percentage by which the number of points in a tile is allowed to **deviate from the expected value** (Tile Accuracy).” See col. 10, lines 3-6. “An additional

accuracy parameter is the **probability** (as a percentage) of a tile satisfying the accuracy criterion. This percentage (TilePercentage) is the number of tiles of the total number that must satisfy the accuracy criterion for the model to be judged acceptable.” See col. 10, lines 6-11.

### Claims 1, 5, and 6

Contrary to Fayyad, claims 1, 5, and 6 recite “calculating a percent of proxy values for the plurality of variables that equals a **mode** of that observation’s corresponding cluster’s proxy values”.

As evidenced by the attached 37 CFR § 1.132 Declaration of Dr. Bo Honore, one skilled in the art would find that Fayyad fails to establish a prima facie case of anticipation because Fayyad does not teach expressly or inherently “calculating a percent of proxy values for the plurality of variables that equals a **mode** of that observation’s corresponding cluster’s proxy values”. According to Dr. Honore, one of ordinary skill in the art would recognize that a **“mode”** is the **“value or item occurring most frequently in a series of observations or statistical data.”** Moreover, Dr. Honore’s Declaration establishes that Fayyad does not teach any “mode” whatsoever. Accordingly, it is respectfully submitted that the rejection of claims 1, 5, and 6 is unsupported by Fayyad and should be reversed.

### Claims 2, 7, and 8

It is well-established that the patent applicant can act as the lexicographer of the patent application, so long as the patent applicant clearly states any special definitions of the claim terms in the patent specification or file history. *Id.* Even when guidance is not provided in explicit definitional format, “the specification may define claim terms ‘by implication’ such that the meaning may be ‘found in or ascertained by a reading of the patent documents.’” *Bell Atl. Network Servs., Inc. v. Covad Communications Group, Inc.*, 262 F.3d 1258, 1268 (Fed. Cir. 2001) (quoting *Vitronics*, 90 F.3d at 1582, 1584 n.6).

Contrary to Fayyad, claims 2, 7, and 8 recite “estimating a **purposeful probability**”. The phrase “purposeful probability” is **defined** in the current application at page 25, line 14 through page 26, line 1. Specifically, “[f]or a question,  $k$  that has  $L_k$  possible answers, the probability (also known as “purposeful probability”) that answer value  $\ell$  is selected by observations (e.g. survey respondents) in segment  $m$  is estimated by

$$\hat{P}_m(k, \ell) = \frac{N_m(k, \ell)}{N_m} (1 - \delta \langle L_k \rangle) + \delta \quad (2)$$

where

$N_m$  = total number of observations in segment  $m$

$N_m(k, \ell)$  = the number of observations in segment  $m$  that gives the  $\ell$ -th answer to question  $k$

$$\delta = \min \left\{ 0.02, \frac{1}{2L} \right\},$$

In the present application, the term “purposeful probability” clearly has been defined in the specification, and that definition must control examination of those claims that recite this term.

As evidenced by the attached 37 CFR § 1.132 Declaration of Dr. Bo Honore, one skilled in the art would find that Fayyad fails to establish a prima facie case of anticipation because Fayyad does not teach expressly or inherently “estimating a **purposeful probability**”. Accordingly, it is respectfully submitted that the rejection of claims 2, 7, and 8 is unsupported by Fayyad and should be reversed.

## VIII. CLAIMS APPENDIX

Appendix A sets forth all pending claims in the state in which they were appealed.

**IX. EVIDENCE APPENDIX**

Appendix B sets forth a copy of the Declaration Under 37 C.F.R. § 1.132 of Dr. Bo Honore previously submitted with the Reply Under 37 C.F.R. § 1.116 on 2 July 2004, which was entered into the record via the Advisory Action dated 30 August 2004.

**X. RELATED PROCEEDINGS APPENDIX**

No decisions have been rendered by a court or the Board in any related proceeding.

**SUMMARY**

In view of the above, Applicants submit that all claims on appeal distinguish over the cited art and respectfully request that the Examiner's rejections of these claims should be reversed.


Applicants therefore respectfully request that the Board of Patent Appeals and Interferences reverse the Examiner's decision rejecting claims 1, 2, and 5-8 and direct the Examiner to pass the case to issue.

The Office is hereby authorized to charge any additional fees or credit any overpayments under 37 C.F.R. §1.16 or §1.17 to Deposit Account No. 50-2504. The Examiner is invited to contact the undersigned at 434-972-9988 to discuss any matter regarding this application.

Respectfully submitted,

Michael Haynes PLC

Date: 3 December 2004

  
\_\_\_\_\_  
Michael N. Haynes  
Registration No. 40,014

1341 Huntersfield Close  
Keswick, VA 22947  
Telephone: 434-972-9988  
Facsimile: 815-550-8850





**PATENT**

**Serial No. 09/867,803**

**Attorney Docket No. 1005-006**

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) : Lawrence J. Choi et al.  
Serial No. : 09/867,803  
Filed : 31 May 2001  
For : METHOD AND SYSTEM FOR CLUSTERING  
OPTIMIZATION AND APPLICATIONS  
Art Unit : 2177  
Examiner : Khanh B. Pham

**Mail Stop Appeal Brief-Patents**

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

**APPENDIX A**

1. (Original) A computer-assisted method for evaluating a cluster assignment for an observation, comprising the activities of:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters;

for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and

outputting the percent for each observation.

2. (Original) A computer-assisted method for evaluating a cluster assignment for an observation, comprising the activities of:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and

outputting each purposeful probability.

3. (Previously Presented) The method of claim 1, further comprising the activities of:

for each observation from the plurality of observations in each cluster from the plurality of clusters, calculating a serendipity probability for each possible value, the serendipity probability is a measure of a probability that an observation in a particular cluster will be randomly associated with any one of the plurality of possible values for a particular variable;

for each observation from the plurality of observations, calculating a ratio of the purposeful probability to the serendipity probability;

for each observation from the plurality of observations, calculating a logarithm of the ratio to obtain composition analysis score; and

outputting the composition analysis scores for each observation in each cluster.

4. (Previously presented) The method of claim 1, further comprising the activities of:
  - for each observation from the plurality of observations, assuming that before the observation can be made, the observation has an equal probability of being in any identified cluster from the plurality of clusters;
  - for each observation from the plurality of observations, assuming that the purposeful probabilities are true;
  - for each observation from the plurality of observations, using Bayes' Theorem to calculate a Bayes probability that a particular observation can be in each cluster conditional upon the observation's proxy value to each variable;
  - outputting the Bayes probability that each observation can be in each cluster.
5. (Original) A computer-readable medium containing instructions for activities comprising:
  - for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters;
  - for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and
  - outputting the percent for each observation.
6. (Original) An apparatus for evaluating a cluster assignment for an observation, comprising:
  - for each of a plurality of observations, means for obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters;

for each observation from the plurality of observations, means for calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and  
means for outputting the percent for each observation.

7. (Original) A computer-readable medium containing instructions for activities comprising:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and

outputting each purposeful probability.

8. (Original) An apparatus for evaluating a cluster assignment for an observation, comprising:

for each of a plurality of observations, means for obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, means for estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and

means for outputting each purposeful probability.



**PATENT**

**Serial No. 09/867,803**

**Attorney Docket No. 1005-006**

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) : Lawrence J. Choi et al.  
Serial No. : 09/867,803  
Filed : 31 May 2001  
For : METHOD AND SYSTEM FOR CLUSTERING  
OPTIMIZATION AND APPLICATIONS  
Art Unit : 2177  
Examiner : Khanh B. Pham

**Mail Stop Appeal Brief-Patents**

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

**APPENDIX B**

Attached herewith is a copy of the Declaration Under 37 C.F.R. § 1.132 of Dr. Bo  
Honore previously submitted with the Reply Under 37 C.F.R. § 1.116 on 2 July 2004, which  
was entered into the record via the Advisory Action dated 30 August 2004.



**PATENT**

**Serial No. 09/867,803**

**Attorney Docket No. 1005-006**

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicant(s) : Lawrence J. Choi et al.  
Serial No. : 09/867,803  
Filed : 31 May 2001  
For : METHOD AND SYSTEM FOR CLUSTERING  
OPTIMIZATION AND APPLICATIONS  
Art Unit : 2177  
Examiner : Khanh B. Pham

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**DECLARATION UNDER 37 C.F.R. § 1.132**

Sir:

I, Dr. Bo Honore, a citizen of Denmark and permanent resident of the United States,  
whose full post office address is Department of Economics, Princeton University, Princeton, NJ  
08544-1021, declare as follows under penalty of perjury:

1. I hold a Ph.D. degree in Economics the University of Chicago awarded in 1987.
2. I am currently a Professor of Economics at Princeton University.
3. I am the Director of the Gregory C. Chow Econometric Research Program at  
Princeton University.

4. I have taught at Northwestern University, and have held Visiting Positions at the University of Chicago and the University of Copenhagen.
5. Since 1987, I have worked continually in the field of econometrics.
6. During my career, I have served on the editorial boards of *Econometrica*, *Journal of Econometrics*, *Review of Economic Studies*, *Econometric Theory*, and *Economic Letters*.
7. I am a Fellow of the Econometric Society.
8. I have reviewed U.S. Patent Application Serial No. 09/867,803.
9. I have reviewed U.S. Patent No. 6,633,882 (Fayyad).
10. Among the methods with which I was familiar prior to 31 May 2001, the filing date of Application Serial No. 09/867,803, were methods of the type recited in Fayyad.
11. I have reviewed the U.S. Patent Office Action dated 3 May 2004 relating to Application Serial No. 09/867,803 (the "Office Action"), which contains the following statement: "As per claims 1, 5, 6, Fayyad teaches ... for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables".
12. That statement in the Official Action is factually incorrect in view of the state of the econometrics art as of the filing date of Application Serial No. 09/867,803. One

- skilled in the art would not find that “Fayyad teaches ... for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation’s corresponding cluster’s proxy values for the corresponding variables”.
13. Rather, upon reviewing the entire specification of the 09/867,803 application, one skilled in the art would recognize that the term “mode” refers to the “value or item occurring most frequently in a series of observations or statistical data”. This is a well-known definition, and matches the definition provided in American Heritage College Dictionary, Third Edition.
14. The Office Action contains the following statement: “Regarding claims 1, 5, and 6, .... Fayyad teaches ... the number of database records in each cluster is the ‘mode of that observation’s corresponding cluster’s proxy value’”.
15. That statement in the Official Action is factually incorrect in view of the state of the econometrics art as of the filing date of Application Serial No. 09/867,803. One skilled in the art would not agree that “the number of database records in each cluster is the ‘mode of that observation’s corresponding cluster’s proxy value’”.
16. Rather, based on the well-known definition of “mode”, one skilled in the art would not find that Fayyad teaches a mode in any manner.
17. The Office Action contains the following statement: “Regarding claims 2, 7, 8,



Fayyad teaches ... estimating a purposeful probability”.

18. That statement in the Official Action is factually incorrect in view of the state of the econometrics art as of the filing date of Application Serial No. 09/867,803. One skilled in the art would recognize that phrase “purposeful probability” is defined in the current application at page 25, line 14 through page 26, line 1 of Application Serial No. 09/867,803. Specifically, “[f]or a question,  $k$  that has  $L_k$  possible answers, the probability (also known as “purposeful probability”) that answer value  $\ell$  is selected by observations (e.g. survey respondents) in segment  $m$  is estimated by

$$\hat{P}_m(k, \ell) = \frac{N_m(k, \ell)}{N_m} (1 - \delta \langle L_k \rangle) + \delta \quad (2)$$

where

$N_m$  = total number of observations in segment  $m$

$N_m(k, \ell)$  = the number of observations in segment  $m$  that gives the  $\ell$ -th answer to question  $k$

$$\delta = \min \left\{ 0.02, \frac{1}{2L} \right\}$$

19. The Office Action contains the following statement: “Regarding claims 2, 7, 8, .... at Col. 22 lines 10-20 [sic] Fayyad teaches the probability for each possible value (i.e., ‘FullTime Sallary’ [sic], ‘FullTime Hourly’, ‘Contract’, ‘Part Time’) of a particular variable (i.e., ‘Employment Status’)”.

**PATENT**

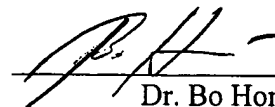
**Serial No. 09/867,803**

**Attorney Docket No. 1005-006**

20. In view of the state of the econometrics art as of the filing date of Application Serial No. 09/867,803, and based upon definition of "purposeful probability" provided in Serial No. 09/867,803, the alleged teaching of Fayyad at col. 22, lines 10-20, or anywhere else in Fayyad, does not show that Fayyad teaches "estimating a purposeful probability".

I further declare that all statements made herein of my own knowledge are true and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code and that willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signed this 30<sup>th</sup> day of June 2004

  
Dr. Bo Honore